

On the Extension of the Mover-Stayer Model when Rate of Transition Follows Negative Binomial Distribution

Adams Y. J. *, Abdulkadir S. S.

Modibbo Adama University of Technology, Yola, Nigeria

Abstract The extension of the Mover-Stayer Model proposed by Blumen, Kogan and McCarthy (1955) is an active area of research. Spilerman (1972) extended the basic model by specifying gamma distribution for the transition rate, the mixture of which resulted in Negative Binomial distribution. However, the Negative Binomial distribution being a unimodal distribution, may not capture situations where excess zeroes exist in the distribution of movements. This paper extends the model using Negative Binomial distribution to model rate of transition in Poisson distribution, which gave the Polya-Aeppli distribution (which is bimodal) as a mixture. The obtained model was validated using a simulated data adopted from Spilerman (1972).

Keywords Mover-Stayer, Polya-Aeppli, Negative Binomial, Bimodal

1. Introduction

Markov chain model has a wide range of applications in social mobility. The model requires assumptions of stationarity and population homogeneity, that is, every element (individual) has the same probability of moving, say, from state i to state j , and the movement (transition) conditioned only on the state in the immediate previous time period. But transition from an origin state hardly conform to this assumption (Spilerman, 1972). However, Blumen, Kogan, and McCarthy, (1955) noted in their interesting study of the movement of workers among various industrial aggregates in US that some individuals simply move more often than, or differently from, others. This idea was also found in the study with intergenerational and intragenerational occupational mobility (Hodge, 1966; Lieberman and Fuguitt, 1967), and with geographical migration (Rogers, 1966; Tarver and Gurley, 1965). This principle led to introduction of heterogeneity to the transition (movement) of an individuals in the population during a unit interval. This idea introduces heterogeneity to the transition (movement) of individuals in the population during a unit interval. With the assumption of heterogeneity, all individuals move according to an identical transition when they move but differ in their rate of mobility. Hence this has resulted to development of “mover-stayer” model (Blumen, Kogan, and McCarthy,

1955, Spilerman, 1972). This implies two types of individuals: the stayer, who with probability one remains in the same category during the entire period of study and the ‘mover’ whose changes in category overtime can be described by a Markov chain with constant probability matrix. The model is appropriate for the analysis of geographical migration or intragenerational occupation mobility where repeated moves can be made by a person. Spilerman (1972) observed that most of mobility data lacks significant detail at the individual level, the mover-stayer model can be applied where the construction of sub-population transition matrices is not possible. Goodman (1961) noted that the transition probability matrix for movers, and the proportion of stayers among the individuals in each category at, say, the initial point in time, are unknown; the estimators provided by BKM for the stayer-movers are inconsistent. There is a lot of estimation methods that have been provided in literatures (Goodman, 1960, Liu and Chen, 2015; Morgan, Aneshensel, and Clark, 1983). Frydman (1984) obtained the maximum likelihood estimation of the mover-stayer model’s parameters by direct maximization of the likelihood, while Fuchs and greenhouse (1988) provided expected –maximization (EM).

This paper is motivated by the statement made in Spilerman (1972) that the originators of mover-stayer model, Blumen, Kogan, and McCarthy (BKM), discuss strategies for extending the mover-stayer model to incorporate a wider range of heterogeneity in the rate of transition, but they do not develop such generalization. BKM proposed in their method of generalization of mover-stayer that instead of postulating two types of persons, we should extent to a process which handle several types. They argue that

* Corresponding author:

adamyusjib@gmail.com (Adams Y. J.)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2018 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

relaxation of fixed number of movement assumption does not prevent the population process-level from being Markovian and cited the example where transitions are Poisson events and the population is homogeneous in its transition rate, the Markov requirement will be appropriate. It is on this basis that Spilerman (1972) extend the basic model by specify a distribution for the transition rate to follow gamma distribution. That is, he assumes Poisson process transition and its rate to follow gamma distribution which resulted in Negative binomial distribution. His reasoning for the choice is that there is little prior knowledge about its (i.e. rate) distribution. However, gamma is unimodal and may not capture situation where the mobility data consist of multi-modal. In literature different distributions have been suggested to model the rate of movement (or random effect). For instance, O'Keefe *et.al* (2012) in their study of psoriatic arthritis consider gamma, inverse Gaussian(IG), and Compound Poisson distribution(CP) for random effects in mover-stayer multistate models. Yiu, Farewell and Tom (2016) explore the existence of a stayer population with mover-stayer counting process model on joint damage, gamma, inverse Gaussian and compound Poisson were considered for random-effects. Cook *et.al* (2002) developed a generalized mover-stayer model for panel data where an individual is allows to move among states according to the underlying Markov process until it encounters one of its absorbing states, where he can no longer move.

2. Methodology

2.1. The Mover – Stayer Model

In the mover-stayer first proposed by Blumen and associates (1955), it is noted that the computations of k-step transition matrices from Markov chain consistently underpredicts the main diagonal elements of the observed k-step matrix. That is,

$$P(1) = \begin{bmatrix} p_{11} & \cdot & \cdot & \cdot & p_{1m} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ p_{m1} & \cdot & \cdot & \cdot & p_{mm} \end{bmatrix}$$

Where $P(1)$ is the observed one- step transition matrix, and $P(k)$

$$P(k) = \begin{bmatrix} p_{11}^{(k)} & \cdot & \cdot & \cdot & p_{1m}^{(k)} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ p_{m1}^{(k)} & \cdot & \cdot & \cdot & p_{mm}^{(k)} \end{bmatrix}$$

is the observed k-step transition matrix. If the $P^*(k)$ is predicted using Markov process, the diagonal elements of the predicted transition matrix will be less than the elements in the observed k-step transition matrix. The reason for the inequality in diagonal elements of the two matrices was attributed to some individuals move more often than others, for each time interval, hence BKM suggested decomposition of the matrix into two subpopulations: the movers and the stayers, because some persons are less apt to move than others in each time interval,

$$P(1) = S + (I - S)M \quad (1)$$

as one-step transition where S is a diagonal matrix, the proportion of stayer, $(I - S)$ is also a diagonal matrix, the proportion of persons with potential mobility, and M is the transition matrix for mobile individuals.

$$P^*(k) = S + (I - S)M^k \quad (2)$$

The k-step matrix is given in equation (2).

It was at this point that BKM (1955) postulate more than two subpopulations in which “instead of requiring every person to make a fixed number of transitions in each time interval, we assume that transitions are random occurrences.” Then a Poisson process was assumed for individuals ‘move with designated parameter, λ . This is given by

$$P_v(t) = \sum_{v=0}^{\infty} \frac{(\lambda t)^v e^{-\lambda t}}{v!} M^v \quad (3a)$$

where $P_v(t)$ is the transition matrix in the interval $(0, t)$. For $t=1$ equation (3a) becomes

$$\sum_{v=0}^{\infty} r_v(1) M^v$$

Where

$$r_v(1) = \frac{\lambda^v e^{-\lambda}}{v!} \quad (3b)$$

which is the Poisson probability ($t=1$).

If we assume many subpopulations, say g with different rates of mobility,

The equation (3b) becomes

$$r_v(t) = \sum_{i=1}^g \pi_i \frac{(\lambda_i t)^v e^{-\lambda_i t}}{v!} \quad (4)$$

Where π_i is the proportion of the i^{th} subpopulation who move with rate λ_i , and if the sampling is made from a continuous distribution $f(\lambda)$, we obtain

$$r_v(t) = \int_0^{\infty} \frac{(\lambda t)^v e^{-\lambda t}}{v!} f(\lambda) d\lambda \quad (5)$$

Blumen, Kogan, and McCarthy (1955) developed equation (5). Thereafter the model was extended by

Spilerman (1972) by assuming gamma density for $f(\lambda)$ to obtain negative binomial distribution. That is,

$$r_v(t) = \binom{\alpha + v - 1}{v} \left(\frac{t}{\beta + t} \right)^v \left(\frac{\beta}{\beta + t} \right)^\alpha \quad (6)$$

The interpretation of this is that proportion of the population making v moves in interval $(0, t)$ will satisfy a negative binomial distribution. Beside this extension authors have proposed various generalization of the model (Cook *et al.*, 2002).

We are motivated by the work of Spilerman (1972) and Johnson, Kotz and Kemp (1992) by assuming negative binomial for π_i . The choice of the negative binomial as a mixing distribution is informed by considering the number of transitions required to achieved desire events

$$\pi_i = f(\lambda) = \binom{k-1}{v-1} (1-\rho)^v \rho^{k-v}; \quad K = 1, 2, \dots \quad (7)$$

$$\begin{aligned} r_v(t) &= \sum_{v=1}^{\infty} \frac{(\lambda t)^v e^{-\lambda t}}{v!} \binom{k-1}{v-1} (1-\rho)^v \rho^{k-v} \\ &= e^{-\lambda t} \sum_{v=1}^{\infty} \frac{\{\lambda t(1-\rho)\}^v}{v!} \binom{k-1}{v-1} \rho^{k-v} \end{aligned} \quad (8)$$

$$r_v(t) = e^{-\lambda t} \sum_{v=1}^{\infty} \frac{\{\lambda t(1-\rho)\}^v}{v!} \binom{k-1}{v-1} \rho^{k-v}; \quad k = 1, 2, 3, \dots \quad (9)$$

The equation (9) coincides with Polya-Aeppli distribution, where the sum stops for $v > k$. If the parameter $\rho = 0$, the distribution in (9) reduces to the classical homogenous Poisson distribution (Minkova, 2002; Minkova, 2004; Chukova and Minkova, 2012). The Polya-Aeppli distribution often called the Inflated-Parameter Poisson [IPo(λ, ρ)] is defined as follows (Minkova & Balakrishnan, 2014). We define the number of transitions made before the number of required event as $N(t)$ in the interval $(0, t)$, where $N(t)$ is

$$P(N(t) = m) = \begin{cases} e^{-t\lambda} & m = 0 \\ e^{-t\lambda} \sum_{i=1}^m \binom{m-1}{i-1} \frac{[\lambda t(1-\rho)]^i}{i!} \rho^{m-i} & m = 1, 2, \dots \end{cases} \quad (10)$$

The mean and variance of the IPo(λ, ρ) distribution are given by;

$$E(N) = \frac{\lambda}{1-\rho} \quad \text{and} \quad Var(N) = \frac{\lambda(1+\rho)}{(1-\rho)^2} \quad (11)$$

The Fisher Index of dispersion is $\frac{Var(N)}{E(N)} = \frac{1+\rho}{1-\rho} > 1$.

Therefore, not only Poisson process is a particular case of Pólya-Aeppli process, but for $\rho \neq 0$, the Pólya-Aeppli process is over-dispersed, which provides a greater flexibility in modeling count data than the standard Poisson process (Chukova & Minkova, 2012).

In general, the Neyman Type A and Thomas distributions can have any number of modes from one upwards. The Pólya-Aeppli distribution, on the other hand, has either one or two modes, while the negative binomial has always one mode (Ascombe, 1950).

The most commonly used distribution to model overdispersed data is the negative binomial, but other distributions may be more appropriate for modelling data with excess zeros, because, unlike the negative binomial, they can have more than one mode, including a mode at zero. Examples include the Neyman Type A and Pólya-Aeppli distributions (Ridout *et al.*, 1998).

3. Parameter Estimation and Model Validation

3.1. Moment Estimates of λt and ρ of the Polya-Aeppli Distribution

$$E(X) = \bar{X} = \frac{\lambda t}{1-\rho} \quad (12)$$

$$Var(X) = \sigma^2 = \frac{\lambda t(1+\rho)}{(1-\rho)^2} \quad (13)$$

From (11) and (12) above,

$$\lambda t = \bar{X} (1-\rho) \quad (14)$$

$$\lambda t = \frac{\sigma^2 (1-\rho)^2}{1+\rho} \quad (15)$$

$$\therefore \bar{X} (1-\rho) = \frac{\sigma^2 (1-\rho)^2}{1+\rho}$$

$$\bar{X} = \frac{\sigma^2 (1-\rho)}{1+\rho}$$

$$\bar{X} (1+\rho) = \sigma^2 (1-\rho) \quad (16)$$

Equation (16) simplified for ρ , becomes

$$\hat{\rho} = \frac{\sigma^2 - \bar{X}}{\sigma^2 + \bar{X}} \quad (17)$$

Substituting (17) in (14) gives, $\hat{\lambda} t = \frac{2\bar{X}^2}{\sigma^2 + \bar{X}}$.

Therefore, the remaining parameters of the model λ and ρ of the Polya Aeppli distribution, can be estimated directly from observed data on the number of moves by an

individual. If \bar{v} and σ_v^2 are the sampling mean and variance of this variable (number of moves), then estimates of λ and ρ can be obtained in terms of these values (see also Minkova, 2012, p. 49). This yields

$$\hat{\rho} = \frac{\sigma_v^2 - \bar{v}}{\sigma_v^2 + \bar{v}} \quad \text{and} \quad \hat{\lambda}t = \frac{2\bar{v}^2}{\sigma_v^2 + \bar{v}} \quad (18)$$

Spilerman (1972), estimated α and β , the parameters of the negative binomial distribution using;

$$\hat{\beta} = \frac{\bar{v}}{s^2 - \bar{v}} \quad \text{and} \quad \hat{\alpha} = \hat{\beta}\bar{v}; \text{ from the observed data.}$$

3.2. Testing the Model with Simulated Data

In order to validate the proposed model, we adopt the structure of the simulated data as provided by Spilerman, 1972 (p. 610), where in the absence of full knowledge of the actual mobility characteristics of the hypothetical population,

individual level transition matrix and a population distribution by rate of movement is presented by assuming six types of persons in the population who move in accordance with Poisson process specified by $\lambda = 0.1, 1.0, 2.0, 3.0, 4.0$ and 5.0 , while the states of the process defined by four geographic regions, given rise to a 4×4 transition matrix (Table 1). The Poisson distribution was used to generate an expected proportion of each subpopulation who make $v = 0, 1, 2, \dots$ moves during the time interval $(0,1)$. These values, multiplied by their respective subpopulation proportions in the total population, were aggregated to produce a distribution of the total population by number of moves. In this case the Poisson estimates were considered as the observed data, and the expected frequencies generated using negative binomial and Polya – Aeppli distributions, and comparison were made among the three distributions using Chi-square as Goodness of fit test. Out of the three distribution, Polya-Aeppli has the smallest chi-square.

Table 1. Structure of the simulated data

A. INDIVIDUAL LEVEL TRANSITION MATRIX	B. DISTRIBUTION OF THE POPULATION BY RATE OF MOBILITY	
	λ	Proportion of the Population with this λ value
$M = \begin{bmatrix} 0.600 & 0.200 & 0.100 & 0.100 \\ 0.150 & 0.700 & 0.100 & 0.050 \\ 0.100 & 0.100 & 0.750 & 0.050 \\ 0.050 & 0.050 & 0.100 & 0.800 \end{bmatrix}$	0.1	0.25
	1.0	0.35
	2.0	0.20
	3.0	0.10
	4.0	0.06
	5.0	0.04
		1.00

Table 2. Distribution of number of moves from observed (simulated) data, negative binomial and from proposed polya aeppli estimates

V	[1000r _v (1)]	[1000f _v (1)]	[1000f _v (1)]
Number of Moves	Number of Persons with v moves (Observed Data)	Number of Persons with v moves (calculated from Negative Binomial; $\alpha = 1.371, \beta = 0.915$)	Number of Persons with v moves (calculated from Polya-Aeppli; $\lambda = 0.9722, \rho = 0.3527$)
	1	2	3
0	388	363	378
1	226	260	238
2	153	161	159
3	97	94	98
4	59	54	57
5	34	30	32
6	19	17	18
7	11	9	9
8	6	5	5
9	2	3	3
10	1	1	1
1000 $\sum r_v(1)$	996*	997*	998*
Chi-Square Statistics (χ^2)		8.34	1.80

Note: $\bar{v} = 1.502$; $\sigma_v^2 = 3.139$

*Value less than 1,000 because of rounding error.

4. Discussion and Conclusions

The Poisson distribution model was used as a baseline model to generate individual level transition matrix and the observed frequency distribution of moves. The expected frequencies were obtained using Polya-Aeppli distribution.

The comparison of the observed frequencies in column 1 of Table 2 with the expected values given in columns 2 and 3 of the same table, shows that Polya-Aeppli distribution fits the observed frequencies much better than the negative binomial distribution which was proposed by Spilerman (1972). The value of χ^2 is 1.80 as against tabular value of 14.067 which shows that the approximation by Polya-Aeppli distribution is acceptable.

In this work, we generalized the Mover-Stayer model by assuming individuals move in accordance with a Poisson process, and that the Negative Binomial density provides a reasonable approximation to the distribution of mobility rates in the population. Moreover, the combination of Poisson distribution and negative binomial density which resulted in the Polya-Aeppli distribution, a bimodal distribution, can conveniently accommodate excess zeroes which was not possible in Spilerman's (1972) extension model. Therefore, the weaknesses of the Spilerman's extension identified by Fang (2013), Rodriguez (2013) and He *et.al* (2014), have been addressed by the obtained Polya-Aeppli distribution. It has also been established that Polya-Aeppli can be used in mover-stayer model.

REFERENCES

- [1] Abdulkadir, S.S. and Mijinyawa, M. (2014). "Application of Mover-Stayer Model to Determine Effective Number of Political Parties in Nigeria." *Bagale Journal of Pure and Applied Sciences*, Vol. 9, No. 2, pp. 32-44.
- [2] Ascombe, F.J (1950). "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions." *Biometrika*, Vol. 37, No. 3/4, pp. 358 – 382. <http://www.jstor.org/>.
- [3] Blumen, I., Kogan, M., and McCarthy, P.J. (1955). *The Industrial Mobility of Labour as a Probability Process*. In Spilerman, Seymour (1972). "Extensions of the Mover-Stayer Model." *American Journal of Sociology (AJS)*, Vol. 78, No. 3, pp. 599-626.
- [4] Chukova, S. and Minkova, L.D. (2012). "Characterization of the Polya-Aeppli process." [Electronic version]. Retrieved from, Sms.victoria.ac.nz/foswiki/pub/Main/ResearchReportSeries/MSOR12-02.
- [5] Cook, R.J., Kalbfleisch, J.D and Yi, G.Y (2002). "A generalized mover-stayer model for panel data." *Biostatistics*, Vol. 3, No. 3, pp. 407-420. Retrieved June 11, 2015 from <http://biostatistics.oxfordjournals.org>.
- [6] Fang, R. (2013). "Zero-Inflated Negative Binomial Regression Model for Over-Dispersed Count Data with Excess Zeros and Repeated Measures, an Application to Human Microbiota Sequence Data" [Electronic version]. Master of Science Thesis, University of Colorado.
- [7] Frydman, H (1984). "Maximum Likelihood Estimation in the Mover-Stayer Model." *Journal of American Statistical Association*, vol.79, No. 387, p. 632-638.
- [8] Fuchs, C. and Greenhouse, J.B. (1988). "The EM Algorithm for Maximum Likelihood Estimation in the Mover-Stayer Models." *Biometrics*, 44, 605-613.
- [9] Goodman, L.A. (1961). "Statistical Methods for the Mover-Stayer model." *Journal of American Statistical Association*, 56, 841 – 868.
- [10] Goodman, L.A. (2013). "Generalizations of the Simple Mover-Stayer Model, Association Models and Quasi-symmetric Models, in the Analysis of Brand-switching Data and other cross-classified Data." *Journal of Empirical Generalizations in Marketing Science*, Vol. 14, No. 1, retrieved on October 30, 2016 from; <https://www.empgens.com/wp-content/uploads/2013/06/goodman.pdf>.
- [11] He, H., Tang, W., Wang, W. and Crits-Christoph, P. (2014). "Structural zeros and zero-inflated models." *Shanghai Archives of Psychiatry*, 26(4), 236-242. Retrieved from, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194007/> on January 16, 2017.
- [12] Hoel, P.G., Port, S.C., and Stone, C.J. (1972). *Introduction to Stochastic Processes*. Boston: Houghton Mifflin Company.
- [13] Hodge, R.W. (1966). "Occupational Mobility as a Probability Process" *Demography*, 3: 19 – 34 in Spilerman, S. (1972).
- [14] Hougaard, P., Lee, M.T and Whitmore, G.A. (1997). "Analysis of Overdispersed Count Data by Mixture of Poisson Variable and Poisson Processes" *Biometrics*, Vol. 53, pp. 1225-1238.
- [15] Johnson, N.L., Kotz, S. and Kemp, A.W. (1992). *Univariate Discrete Distributions*, 2nd ed., Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York.
- [16] Lieberman, S. and Fuguitt, G.V. (1967). "Negro-White Occupational Differences in the Absence of Discrimination" *American Journal of Sociology* 73: 188 – 200, in Spilerman, S. (1972).
- [17] Liu, H. and Chen, S. (2015). "Credit Risk Measurement Based on the Markov Chain" *Business and Management Research*, Vol. 4, No. 3, pp. 32-42.
- [18] Minkova, L.D. (2002). "A Generalization of the Classical Discrete Distributions" *Communications in Statistics – Theory and Methods*, Vol. 31(6), pp. 871-888.
- [19] Minkova, L.D. (2004). "The Polya-Aeppli Process and Ruin Problems" *Journal of Applied Mathematics and Stochastic Analysis*, Vol. 3, pp. 221-234.
- [20] Minkova, L.D. (2012). "Distributions in Insurance Risk Models" Unpublished doctoral dissertation, Sofia University "St.Kl.Ohridski" Bulgaria.
- [21] Minkova, L.D. and Balakrishnan, N. (2014). "On a Bivariate Polya-Aeppli Distribution." *Communications in Statistics – Theory and Methods*, Vol. 43, pp. 5026 – 5038. doi: 10.1080/03610926.2012.709906.
- [22] Morgan, T.M., Aneshensel, C.S. and Clark, V.A. (1983).

- “Parameter Estimation for Mover-Stayer models: analyzing depression over time.” *Sociological Methods and Research*. 11: 345-366.
- [23] O’ Keefe, A. G., Tom, B.D.M and Farewell, V.T. (2012). “Mixture distributions in multi-state modelling: some considerations in a study of psoriatic arthritis, *Statist. Med.*, 32, 600 – 619. Willey Online Library in Yiu, S., Farewell, V.T. and Tom, B.D.M. (2016).
- [24] Ridout, M., Demetrio, C.G.B. and Hinde, J. (1998). “Models for count data with many zeros.” [Electronic version]. A paper presented at the International Biometric Conference, Cape Town, December 1998.
- [25] Rodriguez, G. (2013). Models for Count Data with Overdispersion [Electronic version]. Retrieved from Data.princeton.edu/wws509/notes/c4a.
- [26] Schlattmann, P. (2009). “Medical Application of finite Mixture Models” Springer, p. 33.
- [27] Spilerman, S. (1970). “Extensions of the Mover-Stayer Model.” Institute for Research on Poverty. Discussion Papers, University of Wisconsin http://www.irp.wisc.edu/publications/dps/pdfs/dt_7970.pdf.
- [28] Spilerman, S. (1972). “Extensions of the Mover-Stayer Model.” *American Journal of Sociology (AJS)*, Vol. 78, No. 3, pp. 599-626, <http://www.jstor.org/stable/2776309>.
- [29] Yiu, S., Farewell, V.T. and Tom, B.D.M. (2016). “Exploring the Existence of a Stayer Population with Mover-Stayer Counting Process Models: application to joint damage in psoriatic arthritis.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. [Electronic version], printed on January 10, 2017, <http://onlinelibrary.wiley.com/doi/10.1111/rccs.12187>.